DOCUMENT RESUME

ED 141 371 TM 005 773

AUTHOR Lee, Lai-Min Paul; Coffman, William E.
TITLE A Study of the I Don't Know Response in Multiple-Choice Tests. Iowa Testing Programs Occasional Papers, No. 5.
INSTITUTION Iowa Testing Programs, Iowa City.
PUB DATE Mar 74
NOTE 20p.
AVAILABLE FROM Iowa Testing Programs, 334 Lindquist Center for Measurement, University of Iowa, Iowa City, Iowa 52242 ($0.25)

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Achievement Tests; Elementary Education; Grade 5; Grade 8; *Guessing (Tests); Item Analysis; Mathematics; *Multiple Choice Tests; Reading Comprehension; *Response Mode; Sex Differences; *Standardized Tests; Statistical Analysis; Student Characteristics; Test Construction; *Test Items; Test Results; Vocabulary
IDENTIFIERS *Dont Know Responses; Iowa Tests of Basic Skills

ABSTRACT

Multiple-choice standardized achievement tests of English vocabulary and reading comprehension and of mathematics were administered to samples of 592 grade eight students and 615 grade five students. Two forms of each test unit were prepared. The control groups took forms containing items with four responses, while the experimental groups took forms which had an additional response of I don't know. A few fictitious vocabulary items having no right answers were included in each of the English test units. In grade eight tests the mean scores of the control groups were higher. For the grade five samples there were no differences in mathematics; and differences in English (the control group obtaining higher scores) were found only for low ability students. Item discrimination indices obtained from the two forms did not show any significant differences. There was a negative linear relationship between percentage choosing the I don't know response and percentage correct. In general, lower ability students used the response more often than those with higher ability except in the case of the fictitious items. Sex differences were also found to be a factor in the use of the I don't know response. (Author/MV)

IOWA TESTING PROGRAMS OCCASIONAL PAPERS

Number 5 - March, 1974

A STUDY OF THE *I DON'T KNOW* RESPONSE IN MULTIPLE-CHOICE TESTS

Lai-Min Paul Lee and William E. Coffman

University of Iowa

Abstract

Multiple-choice standardized achievement tests of English vocabulary and reading comprehension and of mathematics were administered to samples of 592 grade eight students and 615 grade five students. Two forms of each test unit were prepared. The control groups took forms containing items with four responses, while the experimental groups took forms which had an additional response of *I don't know*. A few fictitious vocabulary items having no right answers were included in each of the English test units.

It was found that in grade eight tests the mean scores of the control groups were higher. For the grade five samples there were no differences in mathematics; and differences in English (the control group obtaining higher scores) were found only for low ability students. Item discrimination indices obtained from the two forms did not show any significant differences. There was a negative linear relationship between percentage choosing the *I don't know* response and percentage correct. Some characteristics of students choosing the *I don't know* response were identified.

A STUDY OF THE *I DON'T KNOW* RESPONSE IN MULTIPLE-CHOICE TESTS

Lai-Min Paul Lee and William E. Coffman

Although multiple-choice testing has been accepted and extensively used in determining achievement in schools and in making decisions about college admission and hiring, one of the fundamental issues in this form of testing is the problem of guessing. The tendency to guess by examinees appears to be not a stable trait, but rather one that may vary with the age, sex, race, personality, or motivation of the examinees (Votaw, 1936; Swineford, 1938, 1941; Gritten & Johnson, 1941; Sheriffs & Boomer, 1954; Slakter, 1967). Various researchers have examined the effect on item difficulties of response alternatives designed to discourage guessing (Wesman & Bennett, 1946; Rimland, 1960; Williamson, 1967). To take account of the guessing factor in multiple-choice testing, a number of correction formulae have been proposed (Guilford, 1936; Horst, 1933). Recently Gene Glass (Burton, 1972) adapted a correction formula to situations in which examinees could eliminate more than zero but less than a-1 of the incorrect alternatives.

There are not many reports in literature on the use of *I don't know* as an alternative in multiple-choice tests to reduce guessing (The Secondary School Examination Council, 1964; Knapp, 1968; Burton, 1972). The Metropolitan Achievement Tests and a large number of National Assessment Exercises include *I don't know* as one of the options. Recent National Assessment Science Exercises results (Sherman, 1973)

indicated that the tendency to show *I don't know* response increased with age, while the tendency to choose incorrect alternatives decreased with age. In general, females at all ages gave *I don't know* responses to multiple-choice science exercises more often than males. Blacks tended to use *I don't know* responses more often than the nation as a whole, even though the differences may be negligible at the three younger age levels.

The present study was designed to investigate the effects of an additional response of *I don't know* to items in multiple-choice standardized achievement tests of English vocabulary and reading comprehension and of mathematics in grades five and eight. Two forms of each test unit were prepared. The forms containing items with four responses were given to groups of subjects referred to as the control groups. The experimental groups received the forms of test units containing identical items except with an additional *I don't know* response.

The following questions were posed and studied in this investigation:

1. Will the scores of the control groups on tests with four alternative responses be higher than the scores of the experimental groups on corresponding tests with an additional *I don't know* response?

2. Are there any particular items with four responses that differ significantly in item difficulty from the same items with an additional *I don't know* response?

3. How would the items with an additional *I don't know* response compare in item discrimination with the items without this response?

4. What is the relationship, if any, between the percentage choosing the *I don't know* response and the percentage getting the item correct in tests containing this additional response?

5. What are the characteristics of students who choose the *I don't know* response on tests with this additional response?

Procedure

A total of 592 grade eight students and 615 grade five students from eighteen schools in eleven school districts in the State of Iowa were included in the study. Four tests were constructed for each grade, two in English and two in mathematics. The test units for the control groups contained items with four responses. The test units for the experimental groups were identical except that each item contained a fifth response, *I don't know*. Items in these units were selected from various forms of Iowa Tests of Basic Skills (ITBS). Four fictitious vocabulary items with no correct answers were included in each of the English tests. The testing time for each test was 20 minutes. The four tests units were distributed so that each successive student in the class received a different test, resulting in approximately equivalent fourths of the students taking each test unit. All these students had taken the 1973 regular examination of ITBS; scores from the appropriate ITBS subtests were used as the criterion scores

for analyzing the items in the experimental tests and for controlling for random differences in the basic ability between experimental and control groups.

Analysis

Item difficulty (p-rights) and discrimination indices (r-biserial) were calculated for each item for each of the four experimental tests using scores on the related ITBS subtest as the criterion. The technique of analysis of covariance, using the related ITBS subtest scores as the control, was used to investigate whether scores of the control group were higher than scores of the experimental group. An arc sine transformation of the p-rights was performed to compare the item difficulties of items of the control group and the experimental group. A matched pair t-test was used to compare the r-biserials of items in the tests of the control group and the experimental group. The technique of analysis of covariance was also used to compare percentages choosing the *I don't know* response for the items in each of the four tests containing this additional response. In this case, the control variable was the percentages choosing the right answers for these tests. The characteristics of the group choosing the *I don't know* response was also analyzed by tabulating the percentage of students using this response at least once and the mean number of usage of this response by ability level and by sex.

Results

As shown in Table 1, in the grade eight English and mathematics tests, the mean scores of the control group in tests with four reponses were higher than the mean scores of the experimental group in

Table 1

Summary of Criterion Scores and Test Scores

	X Mean	X S. D.	Y Mean	Y S. D.
Grade 5 English				
Control	110.5	28.7	15.3	4.7
Experimental	112.7	27.9	14.9	5.5
Grade 5 Mathematics				
Control	114.4	25.1	17.7	5.6
Experimental	115.2	25.3	17.7	5.7
Grade 8 English				
Control	172.5	41.1	18.7	6.0
Experimental	168.9	37.2	17.0	5.9
Grade 8 Mathematics				
Control	178.2	38.3	21.1	6.1
Experimental	172.9	35.6	19.3	6.1

X is criterion score

Y is test score

tests with a fifth additional response of *I don't know*. The mean scores of the two groups were not different for the grade five mathematics tests. The regression lines of the fifth grade groups on the English tests had different slopes, with the poorer students tending to get lower test scores when the *I don't know* response was available.

Generally, the item difficulty figures were consistent with the mean scores. There seemed to be no overall difference in p-rights in grade five English and mathematics tests. As a whole, the items with four responses in grade eight English and mathematics tests had higher p-rights than corresponding items with the *I don't know* response. For each comparison, when the average difference over all items is taken into account, the remaining differences can be attributed to sampling error.

Results of the matched pair t-tests of r-biserials between items of the control group and the experimental group, as shown in Table 2, did not show any significant differences, even though the criterion tests did not include the *I don't know* response and thus were more like the tests taken by the control group than the tests taken by the experimental group.

There was a linear relationship between percentage choosing the *I don't know* response and percentage correct. The correlation coefficients ranged from -.56 to -.78. The three regression lines of the grade five English, grade five mathematics and grade eight mathematics tests with percentage choosing the *I don't know* response on percentage correct could be regarded as having a single regression slope (Figure 1). The slope of the grade eight English regression

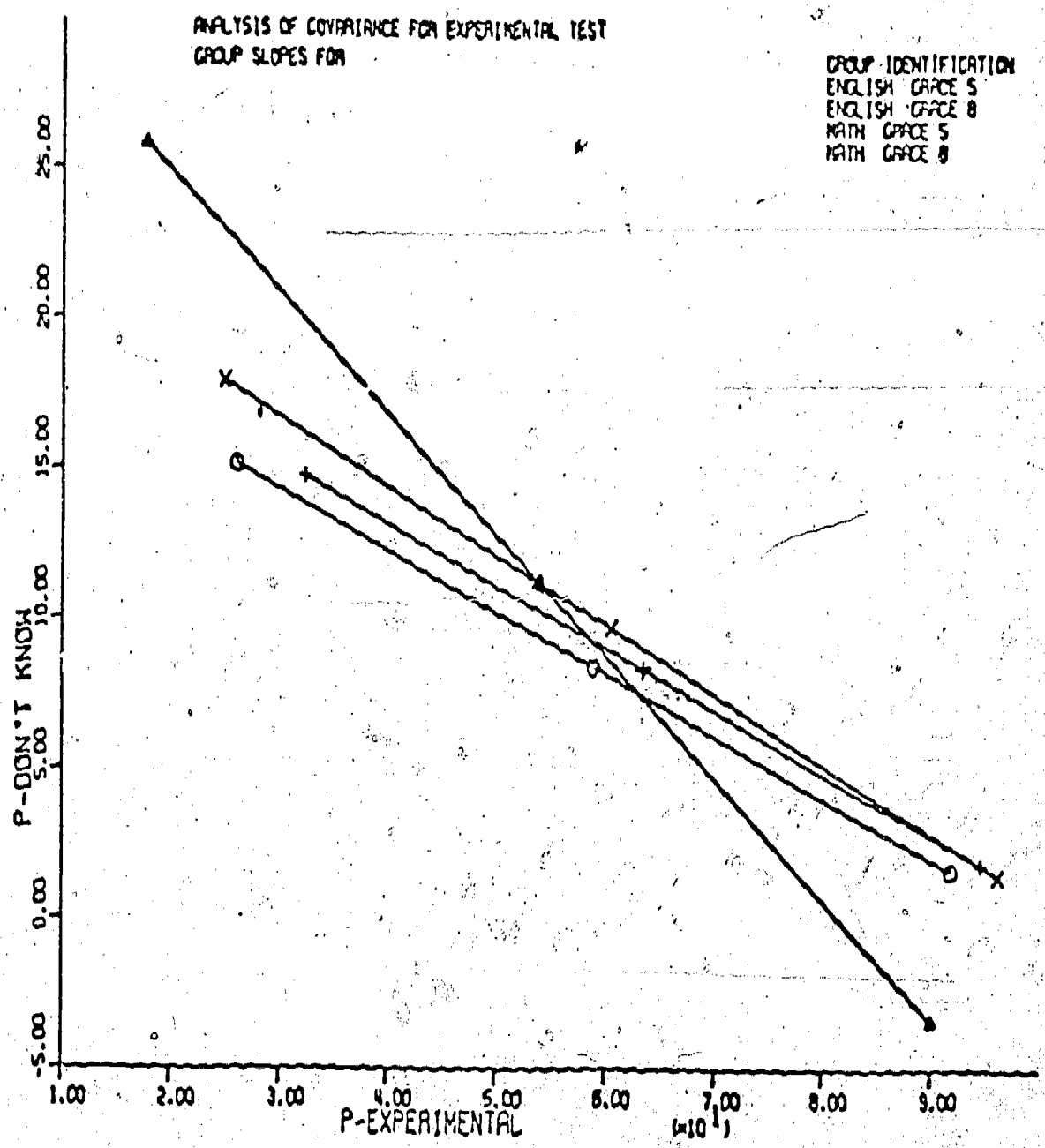
Table 2

Results of Matched Pair t-tests of r-biserial Between Experimental and Control Group Tests

	Mean $r_E - r_C$	S. E. $r_E - r_C$	df	t
Grade 5 English	.05	.025	26	2.00
Grade 5 Mathematics	.00	.019	27	.00
Grade 8 English	.01	.028	31	.36
Grade 8 Mathematics	-.03	.019	35	-1.58

Figure 1

Regression Lines of Percentages Choosing *I Don't Know* Response on P-Rights for the Four Experimental Groups



line, however, was much steeper than the slope of the other regression lines. Results of the comparison of regression showed that more students chose the *I don't know* response in mathematics tests than in English tests; and grade eight students used this response more often than grade five students.

It was observed that most of the students used the *I don't know* response one or more times when given the opportunity. In general, lower ability groups used this response more often than higher ability groups, which is what one would expect if the *I don't know* response were to reduce guessing. As shown in Table 3, much higher percentages of students, across all three ability groups, chose the *I don't know* response in the fictitious items than in the other items, even the very difficult items. In the fictitious items, higher ability students used the *I don't know* response more often than lower ability students.

In grade eight tests, male students used the *I don't know* response more often than female students (Table 4). However, grade eight female students had higher criterion scores in both English and mathematics than male students. In grade five tests, female students chose the *I don't know* response more often than male students, even though female students had higher criterion and test scores than male students. Female students used the *I don't know* response in the fictitious English items more often than male students. This means that in both grades male students had a greater tendency to guess than female students.

Table 3

hoosing *I Don't Know* Response in Grade 5* and Grade 8
English Test by Ability and P-Rights

No. of Items	Ability Group		
	Low	Medium	High
3	*7.16*	*1.26*	*0*
2	10.20	2.00	2.00
5	*14.12*	*3.78*	*.76*
4	7.34	4.08	1.58
4	*11.25*	*6.15*	*3.35*
10	6.33	4.48	4.52
5	*16.48*	*9.06*	*3.44*
2	9.53	8.13	11.10
4	*12.75*	*6.12*	*3.35*
6	16.70	21.24	8.24
4	*21.07*	*17.92*	*8.65*
3	19.76	18.33	15.70
0	-----	-----	-----
5	21.22	30.60	20.78
2	*47.05*	*54.75*	*39.40*
0	-----	-----	-----
Items *4*	*29.40*	*43.40*	*44.23*
4	31.12	49.97	52.95

ıta in italics

Table 4

Average Number of *I Don't Know* Responses Used by Male and Female Students

	Male	Female	Total
Grade 5 English			
(27 real items)	2.98	3.05	3.02
(4 fictitious items)	1.39	1.68	1.56
Grade 5 Mathematics			
(28 items)	1.86	2.82	2.32
Grade 8 English			
(32 real items)	4.28	2.86	3.57
(4 fictitious items)	1.78	1.82	1.79
Grade 8 Mathematics			
(32 items)	3.38	2.69	3.07

Discussion

The problem of guessing in multiple-choice testing is a very subtle problem. In the present study, the inclusion of the *I don't know* response appeared to be only partially successful in reducing guessing. At the grade five level, the scores of the two groups were not much different. At grade eight, the scores of the control group were higher than those of the experimental group. The latter finding is in agreement with Knapp's study on mathematics (1968).

Results of matched pair t-test of r-biserials failed to show any significant difference between tests of the two groups. It should be remembered, however, that the criterion tests used in the item analyses were multiple-choice tests without *I don't know* responses. In the present study, the ITBS scores were used as the criterion scores in the item analysis rather than the test scores, because the ITBS had more items in the test, thus it would provide a more reliable criterion than the much shorter experimental tests. Furthermore, the criterion score would be independent of the items. As a check on the possibility that differences might be greater if the criterion had been total score of which the item was a part, an additional item analysis was carried out. The biserial correlations tended to be about .10 higher (reflecting the dependencies), but the average differences between experimental and control groups remained essentially the same.

The negatively correlated linear relationship between percentage choosing the *I don't know* response and percentage correct is as expected. This finding agrees with the findings by Sherman (1973) for

National Assessment Science exercises. Furthermore, the increase in tendency to use the *I don't know* response with age is also found in Sherman's studies. It seems that older children are more aware of what they don't know and more willing to admit they don't know the answers. However, if the *I don't know* response was not provided, as in the tests taken by the control groups, children at both age levels in the present study just guessed, as demonstrated by the low percentage of *omits* and *not reached*.

It is not surprising to find that students of low ability used the *I don't know* response more often than students of high ability. However, for the fictitious items, more high ability students chose the *I don't know* response. Using the number of responses of *I don't know* in fictitious items as an index of guessing, this negative correlation of achievement with the tendency to guess agrees with Slakter's finding (1967), but not with Swineford's finding (1941). It should be noted, however, that Swineford (1941) showed that the male students had a higher tendency to guess than the female students, a finding confirmed in this study.

It may be argued that since the multiple-choice tests contain a source of inaccuracy due to guessing, recall tests that require examinees to supply the answer rather than to choose the best answer from several alternatives should be used instead. However, tests with recall items have their problems also; they are often ambiguous and always difficult to score objectively and speedily, particularly in large-scale testing programs. Though multiple-choice items may

be contaminated by guessing, they are just too useful to be discarded. The more promising approach would be more research, like the present study, to throw light on the nature of guessing in multiple-choice items and to generate procedures for reducing this nuisance factor in multiple-choice tests.

References

Burton, N. W. *An Investigation of Item-Scoring Formulas Which Take Into Account Random Guessing, Partial Information, and Misinformation.* Unpublished doctoral dissertation, The University of Colorado, 1972, (University Microfilms No. 72-25,149)

Gritten, F., & Johnson, D. M. Individual differences in judging multiple-choice questions. *Journal of Educational Psychology*, 1941, *32*, 423-430.

Guilford, J. P. The determination of item difficulty when chance success is a factor. *Psychometrika*, 1936, *1*, 259-264.

Horst, P. The difficulty of a multiple-choice testing item. *Journal of Educational Psychology*, 1933, *24*, 229-232.

Knapp, T. R. The mathematics study. Unpublished report. Exploratory Committee on Assessing the Process of Education, 1968.

Rimland, B. The effects of varying time limits and of using *right answers not given* in experimental forms of the U. S. Navy arithmetic test. *Educational and Psychological Measurement*, 1960, *20*, 533-538.

Secondary School Examination Council. *Examinations Bulletin No. 7. The Certificate of Secondary Education Experimental Examination: Mathematics 2.* London: Her Majesty's Stationery Office, 1965.

Sherman, S. Personal Communication, 1973.

Sheriffs, A. C., & Boomer, D. C. Who is penalized for guessing? *Journal of Educational Psychology*, 1954, *45*, 81-90.

Slakter, M. J. Risk taking on objective examinations. *American Educational Research Journal*, 1967, *4*, 31-43.

Swineford, F. The measurement of a personality trait. *Journal of Educational Psychology*, 1938, *29*, 295-300.

Swineford, F. Analysis of a personality trait. *Journal of Educational Psychology*, 1941, *32*, 438-444.

Votaw, D. F. The effect of do-not-guess directions upon the validity of true-false or multiple-choice tests. *Journal of Educational Psychology*, 1936, *27*, 698-70[illegible].

Wesman, A. G., & [illegible], C. K. The use of *none of these* as an option in test construction. *Journal of Educational Psychology*, 1946, *37*, 541-549.

Williamson, M. L., & Hopkins, K. D. The use of *none of these* versus homogeneous alternatives on multiple-choice tests: Experimental validity and reliability comparisons. *Journal of Educational Measurement*, 1967, *4*, 53-58.